

**More than Just Language Proficiency:
Designing a Speaking Test for Study Abroad Candidates
Daniel Bates, Asia University**

Abstract

This article describes the thought processes behind the design of a new systemized test to assess the capabilities of students at Asia University wishing to partake on a study abroad program in the United States. The test includes a set of newly created test items that will allow examiners to more easily refer to an already developed rubric to assess not only a candidate's language proficiency skills, but also their intercultural understanding. The framework of the original test is examined, along with other commercial language proficiency tests, before a detailed explanation of the new systemized test is introduced. The reasoning behind the new structure, along with the rationale for using the test items and topics, is presented alongside the existing rubric.

Introduction

This paper begins with a thorough examination of a speaking test used at Asia University as a supplementary assessment for students from the Urban Innovation (U.I) department wanting to partake in the University's study abroad program, but who did not obtain the required score of 600 on the TOEIC language proficiency test. A clear scoring rubric, based upon a distinct theory of holistic assessment, a set of "performance assessments centered around humanistic constructs like intercultural understanding" (Carpenter & Matsugu, 2016, p. 363), is already in place for assessing students in this test. However, the construct of the actual speaking test was very loose with some significant deviations occurring between test takers that question the overall validity and accuracy of the test. Therefore, the assessments committee at Asia University have developed a new format to systemize this speaking test which complements the existing holistic approach to assessing intercultural understanding. This paper will introduce the new construct, along with the reasoning behind the new structure and the items included within, give a detailed explanation of what the new items are testing for, and how they fit in with the existing rubric and overall theme of intercultural awareness.

This paper begins by looking at the background to this test; the original purpose and reason for it, then the construct of the original test is examined along with the rubric and the theory behind that. Next, the new test construct is explained, with detailed reasoning for each of the sections and items included. The paper ends with some thoughts on potential future modifications and how trialing the test can lead to further improvements to the test items, the construct, and the rubric.

Commercial Language Proficiency Tests

As an examiner of a number of commercial language proficiency tests, it was pertinent to begin this process by underlining similarities between these tests that could be incorporated in the new university test construct. As this university speaking test is not strictly a language proficiency test, the specific items used on these commercial tests themselves are of no particular relevance, but the structure and systematic nature of those tests are. Primarily, each test is designed to give all candidates equal opportunities to show their language capabilities in an

allotted time. There are always several parts to these tests, each containing a range of questions or tasks that allow the candidate a number of opportunities to start over with a fresh topic. Timing is another crucial constant with each test administered within strict boundaries. On top of that, a clear rubric is in place that allows all trained examiners to grade candidates to the same standards. The systemized elements found in these tests were missing from the previous incarnation of the speaking test with a lack of continuity in topics discussed and some significant differences in the length of each test.

Test Background

The original speaking test was developed at Asia University as part of the application for students to partake on a study abroad program with a number of partner universities in the United States. This program in the US lasts for six months, and in addition to English language classes, the students also do an internship at a Japanese company based in the US. As such, one of the stipulations required of the students to study abroad was to achieve a TOEIC score of 600 or above. However, in an effort to allow more students to sign up for this program, a second route has been made available. Students with a TOEIC score of at least 400 could take an additional speaking test worth up to 200 points, allowing them to potentially reach the required score of 600. Thus, a speaking test was built for this purpose which encouraged examiners to give a score based upon a student's perceived readiness or otherwise to study, work and interact in an English speaking country.

Original Test Construct

The original speaking test was built with the goal of assessing both a students' English language skills and also their "ability to survive and thrive in the United States as international students" (Carpenter & Matsugu, 2017). This was reflected in the scoring rubric with equal scores (20% each) given to English Ability, Degree of Detail, Participation, Cultural Fit and Comfort. The original test was designed with a clear holistic methodology created for assessment through these rubrics, but little time was taken in the structure and construct of the actual test.

Creating a test requires test constructors to consider and implement a number of things, first, they require a clear outline of who the test taker will be, what the goal of the test is, what content is to be covered and by what methods, how many sections are required as well as the length of the test (Aldersen, Clapham & Hall, 1996). With a test and rubric already in place, an observation and analysis of the original test was undertaken in which three main areas of concern were highlighted; timekeeping, the variety of topics, and discrepancies in aids used by some of the students.

With regards to timekeeping, a rough guide of around ten minutes was set for each candidate. However, with no strict guidelines in place this varied from as little as six minutes to as many as 12 between tests. In terms of topics, some students were able to talk extensively about a topic of their choosing, for example baseball, while other students, possibly more naturally reserved and less outgoing, and thus giving shorter answers, were asked numerous questions on a larger variety of topics. Examiners were asked to commence by asking a set of simple questions about the candidate, but often the students would begin with a self-introduction, some lasting over a minute, that was often memorized. A tester should always have a pattern to follow as stated by Hughes (2003, p. 124) “It is a mistake to begin an interview with no more than a general idea of the course that interview may take”. Finally, some students were referring to notes throughout their test, while others did not. One student even referred to his phone in order to look up some vocabulary. Carpenter & Matsugu (2017, p. 366) stated that “no two interviews were the same in terms of the questions asked”. This level of deviation between each test could call the validity of the test in to question and needed to be standardized in the revised test.

Rubric

The rubric focuses on five areas; English ability, Degree of Detail, Participation, Cultural Fit and Comfort (See Appendix one for the full rubric). Carpenter and Matsugu (2017) go in to significant detail regarding the holistic concept and thought processes behind the creation of the various assessment bands. As such, the new test construct was built with these rubrics already in place. Attempts were made to create test items that would allow examiners to more readily assess these bands in authentic, real life situations. For example, the test rubric includes comfort and

cultural fit, introduced by the original test creator to include spontaneous actions such as handshakes and crossing one's legs while speaking (Carpenter & Matsugu, 2017). The new test tried to further incorporate these concepts by creating more authentic situations that would encourage students to use gestures and spontaneous actions to get their point across through a range of role plays with varying degrees of difficulty.

New Test Construct

The speaking test has far reaching consequences for the students, their families and the university itself, with financial, educational and social implications. With significant differences in the testing circumstance of each test taker, the test in its previous format could make it difficult to justify to those who did not score a sufficient number of points as to why they were not allowed to participate on the study abroad program. As such, when developing the new test, fairness and structure were at the forefront, attempting to create a standardized interview whereby each candidate undertook similar tasks, answered similar questions, talked about similar topics and did these in a comparable amount of time.

The construct and items in this test were devised with a number of things in mind. First, that this is not a language proficiency test per se, but a test of a candidate's suitability for studying abroad, of which their language ability is just a part. As such, the tasks need to provide opportunities for candidates to show their cultural awareness and their comfort levels when using English to complete everyday tasks. Second, as the test takers are all have TOEIC scores of 400-500, the questions and tasks were developed to be understandable and easy to complete for low level English communication. Finally, Hughes (2003) states that a candidate should be given as many fresh starts as possible by using different formats and to be able to interact with more than one examiner and have as many separate items as possible to answer. With this in mind, a three part test was devised, taking on board the aims and requirements of the original test but organizing questions and topics and tasks into three timed sections. The basic construct and the outline of the tasks for the updated speaking test is detailed in the following section (See Appendix A for the full test outline).

Part One (2-3 minutes)

In the previous incarnation of the test, candidates would tend to begin the test with a self-introduction, outlining their name, studies and their hobbies. It could be argued that this went against the intended nature of the test with students being able to rote learn a few basic sentences about themselves, while others were able to speak at length on a familiar topic. In both cases, it is hard to see how either intercultural communication or English abilities could be determined from these responses. Thus, the need to give a self-introduction has been replaced in the updated test format by Part One, a 2-3 minute question and answer session, where examiners are encouraged to ask relatively simple questions about the student's background, likes and dislikes and free time activities. This amendment to the format presents similar topics to the previous test but in an interview style that tests one's ability to listen and interact naturally with the examiners. This part of the test was designed to put candidates at ease, and assess their ability to hold a simple conversation and talk about familiar topics. It was decided to not list any specific questions for examiners to ask in this section, rather just a list of generic topics that would allow for some spontaneity in the communication and reduce the likelihood of students being able to memorize answers beforehand.

Part Two (3 minutes)

After the initial self-introduction, the original test had no set structure, rather just a further range of spontaneous questions, elaborating on the things the candidate had already spoken about or additional questions about culture and life in the United States. These questions related to life abroad and differing cultures in order to help examiners assess the candidate's level of intercultural understanding and awareness, and is something that needed to be retained in the updated version. One of the original requests from the Urban Innovation department was to create a test that resembled a business interview rather than a language test. Therefore, Part Two was designed with a set of predetermined questions that specifically examined the students' attitudes and motivation towards studying abroad, along with their understanding and awareness of different aspects of American life and culture. In particular, examiners could assess for degree of detail and cultural fit in the content of the answers to determine the student's suitability, motivations and willingness to study and live abroad. In order to prevent the possibility of

passing on information to other candidates after finishing the test, two sets of questions based on similar topics and levels of difficulty have been devised.

Part Three (2-3 minutes)

The final part of the test was designed with the original rubrics in mind. By developing a range of simple role plays involving everyday situations such as at a restaurant, at a hospital or giving directions, the candidate would be able to show both their current language abilities and their comfort in trying to convey a message to the examiners. The use of role plays in oral examinations allows for the examinees to ask questions to the examiners, a skill often overlooked in speaking tests (Underhill, 2000, p. 51) and helps create authentic situations that replicate the precise situations the students are likely to encounter while on the study abroad program. The role play section also allows the examiners to adhere closely to each of the five bands in the rubrics, assessing not only a student's English ability, but also their willingness to participate and engage in conversations in English, the amount of detail they give in their responses, whether their demeanor would be appropriate in the US and their comfort while doing so. Three situation cards were developed with a variety of authentic, real-world situations. Item difficulty is also controlled in this section with the situations on each card getting progressively more difficult and featuring more specific and unfamiliar vocabulary as the part progresses. This is to test a student's ability to circumlocute and paraphrase, as making oneself understood clearly despite limitations in one's vocabulary is another necessary skill whilst living abroad. The role plays are based on the following three topics:

1. Ordering food from a menu – An everyday activity that students will have to be able to do effectively from their first day abroad. Role play questions were included to see how students would cope with unfamiliar vocabulary or idiomatic phrases that may be unfamiliar to them.
2. Illness/Ailments – As with the previous topic, this one also features heavily in Freshman English courses that the students will have already taken. Therefore the concept and vocabulary are recognizable, but as the role play continues more difficult scenarios are placed in front of the student to assess if they could successfully get their message across

even if the lexis itself was unknown. Pictures are included so students could attempt to paraphrase and describe what they see and how they might feel in these situations.

3. Directions – A set of destinations are given to the candidate, who then have to give directions based on the map and locations provided. As with the other scenarios, the situations get progressively harder and require more paraphrasing as the test continues.

Examiners

Finally, the examiners themselves, who are already trained to use the rubrics and have experienced examining the original test, will be re-trained to familiarize themselves with the new construct and will be encouraged to speak at a natural, fluent pace. This will allow the candidate the opportunity to ask them to repeat or rephrase the question, and allows the examiners to account for the candidate's comfort or otherwise in doing so. As well as that, testers will be asked to not make notes during the test, and make transitions between questions and sections as natural as possible (Hughes, 2003) in order to reduce the anxieties of the participants.

Conclusion and Recommendations

This test will be implemented for the first time in December 2018. The construct of this test and the rubrics were made independently of each other and although the items were tailored towards the rubrics, there has not been sufficient time to trial this test as yet. The construct of the test and the rubrics used for assessment have been sufficiently developed and discussed in both this and Carpenter and Matsugu's 2017 paper. Therefore, next an extensive trial and statistical analysis is planned which can focus more closely on the test items and the overall relationship between them, the construct and the rubric, allowing for further refinement in the future.

References

- Aldersen, J. Clapham, C. & Wall, D. (1996). *Language Test Construction and Evaluation* Cambridge, Cambridge University Press.
- Bachman, L.F. (1990). *Fundamental Considerations in Language Testing* Oxford, Oxford University Press
- Carpenter, J. & Matsugu, S. (2018). Assessing Intercultural understanding through a speaking test. In P. Clements, A. Krause, & P. Bennett (Eds.), *Language Teaching in a global age: Shaping the classroom, shaping the world*. Tokyo: JALT
- Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.) Cambridge, Cambridge University Press.
- Underhill, N. (2000). *Testing Spoken Language. A handbook of oral testing techniques* Cambridge, Cambridge University Press.

Appendix A

Test Rubric (taken from Carpenter & Matsugu 2017)

	40 points	30 points	20 points	10 points	0 points
English ability	English is clear, and appropriate , and involves minimal listener effort .	English is mostly clear, and appropriate , but involves some listener effort .	English is often unclear , and requires more listener effort .	English is very difficult to understand , and requires a lot of listener effort .	It is very difficult to understand the student because of their language use .
Degree of detail	Answers are thorough, with details and examples .	Answers are mostly thorough , but few details or examples .	Answers are limited , with only one detail or example .	Answers are short , with no details or examples .	Answers do not address the question .
Participation	Student attempts to participate fully in the flow of conversation, and actively engages the examiners .	Student sometimes cannot participate in the flow of conversation, but still tries to actively engage the examiners .	Student often cannot participate in the flow of conversation, and does not actively engage the examiners .	Student only answers the questions , and does not participate in the flow of conversation.	Student does not engage in meaningful conversation.
Cultural fit	Communication style is appropriate for living in the US.	Communication style is usually appropriate for living in the US, although sometimes inappropriate .	Communication style is often not appropriate for living in the US, but overall interaction is still positive .	Communication style is often not appropriate for living in the US, and the overall interaction is negative .	Communication style is not appropriate for living in the US.
Comfort	Comfortable. <ul style="list-style-type: none"> • Eye contact • Facial expression • Body language 	Mostly comfortable	Sometimes uncomfortable, and some anxiety.	Often uncomfortable communicating and much anxiety.	Too nervous to communicate.

Appendix B

Examiner Booklet

Information for examiners –

Please read aloud the sentences **in red** at the beginning of each part.

Examiners should not take notes during the test.

Students should not be allowed to use any notes in the test and cannot consult their phone or a dictionary during the test.

Part 1 – Free questions (2-3 minutes)

Examiner: This test will be in three parts, in this first part we would like to ask you some questions about yourself.

Instructions for examiners:

Examiners are encouraged to ask a variety of spontaneous questions related, but not limited to, the candidates' studies, free time and part time work activities. You should not ask questions about studying/living abroad in this section. Please read ahead to the questions in Part 2 before starting Part 1 to ensure you don't ask questions of a similar nature.

Candidates are not required to introduce themselves. It should be taken as read that they are able to do that – the ability to answer a range of simple questions is being tested here.

Part 2 – Studying and living abroad (3 minutes)

Examiner: In the next section, we will ask you some questions related to living and studying in America.

Instructions for examiners:

Choose question set A or B and ask as many of the questions from that set in the allotted time. If students don't understand the question, please rephrase in a simple and concise manner. If the students give very brief answers without sufficient reasoning or examples, ask a simple follow-up question to elicit further information. If a student completes all the questions and follow up questions before the allotted 3 minutes is complete, move on to Part 3. When the 3 minutes is up, allow the student to finish the sentence and move to part 3 regardless of whether all 5 questions have been completed or not.

In this section, the content of the answer is the most important thing as the questions are directly related to studying abroad. While language use is still important, here we are assessing the candidate's motivations and readiness for living and studying in another country. It will help determine how much thought students have already put into studying and living abroad.

SET A

- 1. What do you like about America?**
- 2. What do you want to do in the future (after you graduate)?**
- 3. What do you think will be difficult about living in the US?**
- 4. What do you hope to gain from AUAP?**

5. How does US culture differ from Japanese culture?

SET B

1. Why do you want to study in the US?
2. What do you think you will miss from Japan while living abroad?
3. What parts of US and Japanese culture are similar?
4. Do you think it will be easy to make American friends when you study abroad?
5. How do you plan to improve your English skills while living in America?

Part 3 – Role play activity (2-3 minutes)

Examiner: In this final section, we are going to do a short roleplay. Please take a look at the card and follow the instructions at the top.

There are 3 role play cards, please choose one per candidate and act out the role play. In this section, improvisation is encouraged from both the examiner and candidate. Feel free to deviate from the role play and ask additional, relevant situational questions. Note, that this section is not only testing the student's language ability in completing these tasks, but also showing their willingness to try and complete the task. Attempts to paraphrase, mime and gesture to get their message across should be looked upon positively as signs of students' willingness to get their message across.

The test is over when either the role play is complete or three minutes have passed.
Choose either:

1. Ordering food from a menu
2. Going to the doctors
3. Giving directions

Role Play 1

I'm hungry – Ordering Food

Students are asked to order the following things from the menu.

- 1) A starter and a sandwich for your main course.
- 2) Ask for a Pizza of your choice and ask for desserts. (ask what kind of pizza they like – types of desserts etc.)
- 3) Ask for more information about the Shepherd's Pie and the Tikka Masala. Choose the one you prefer. (ask if they like spicy food – meat etc.)

Ask simple follow up questions. For example –

Do you want a drink with that?
Can I get you any sides?
What kind of toppings do you want?
How spicy do you want it?

Menu

Starters

Chicken Soup

Salad

Sandwiches -

Ham and cheese

Tuna

Vegetarian

Grilled Cheese

Main Course

Slice of Pizza (Ask for toppings and prices)

Cheeseburger

Shepherd's Pie

Chicken Tikka Masala

Drinks

Coffee

Tea

Soft Drinks - Coke, Sprite, Root Beer, etc.

Role Play 2

I'm sick - Going to the doctors and giving advice

The student will select an illness from each of the first two rows. Imagine you are a doctor and ask some simple questions, such as:

What seems to be the problem?

Ask additional, relevant follow-up questions.

For example:

How long have you (had a cold)?

Have you been taking any medication? /Are you taking anything for it?

Have you been off school?

How did you (break your arm)?

What did you eat?

Do you have any allergies?

Offer some extra advice such as:

Take these pills twice a day

Get some rest

Keep ice on it for 24 hours

Avoid spicy foods

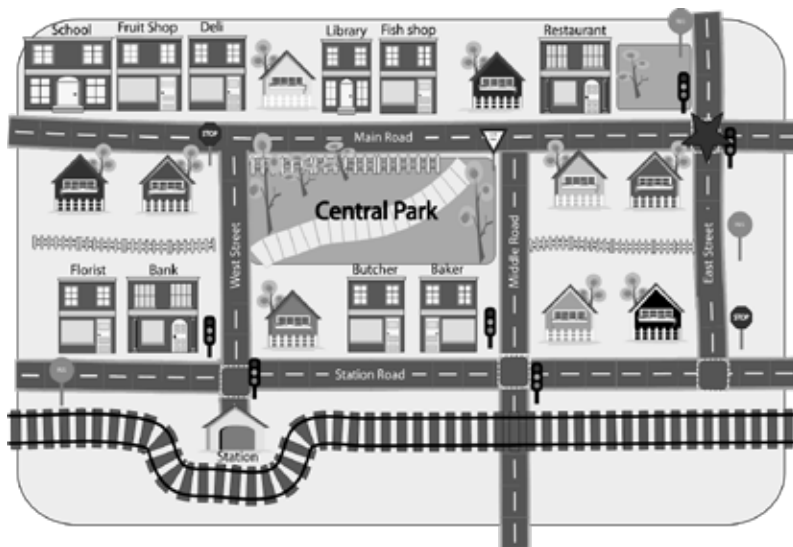
If time allows, switch roles. Describe an illness from the handout and ask the student to give you some advice.

Role Play 2
I'm sick - Going to the doctors

		
		
		
		
have a cold/the flu	have a fever/temperature	have a sore throat
have an earache	have a toothache	feel sick
have a stomach ache	have a headache	have backache
have a broken arm	have sunburn	have a broken leg

Role Play 3

I'm lost - Giving directions



1. I'd like to get to the bank, Could you show me the way?
2. I've gotta pick up some bread and some meat. Do you know where I could do that?
3. I need to pick up my friend from the station. I'd like to take the scenic route and avoid walking on the road. Is it possible?

Continue with additional directions or swap roles if further assessment is deemed necessary.